Inference for Under-Dispersed Data: Assessing the Performance of an Airborne Spacing Algorithm

Sara R. Wilson[1], Robert D. Leonard[2], David J. Edwards[3], Kurt A. Swieringa[1], and Matt Underwood[1]

[1]*National Aeronautics and Space Administration, Hampton, VA*
[2]*Miami University, Oxford, OH*
[3]*Virginia Commonwealth University, Richmond, VA*

ABSTRACT

Poisson regression is a commonly used tool for analyzing rate data; however, the assumption that the mean and variance of a process are equal rarely holds true in practice. When this assumption is violated, a quasi-Poisson distribution can be used to account for the existing over- or under-dispersion. This paper presents an analysis of a study conducted by NASA to assess the performance of a new airborne spacing algorithm. A deterministic computer simulation was conducted to examine the algorithm in various conditions designed to simulate real-life scenarios, and two measures of algorithm performance were modeled using both continuous and categorical factors. Due to the presence of under-dispersion, tests for significance of main effects and two-factor interactions required bias adjustment. This paper presents a comparison of tests of effects for the Poisson and quasi-Poisson models, details of fitting these models using common statistical software packages, and calculation of dispersion tests.

Keywords: Interval Management; Poisson Regression; Quasi-Poisson; Spacing Algorithm; Under-Dispersion.

INTRODUCTION

Many engineering processes require monitoring the number of times an event occurs in a given unit interval. For example, consider a process in which an important quality characteristic is the number of flaws in an object, the number of defective items in a batch, or the number of phone calls per day to a customer service department. These rates are often modeled by a Poisson distribution. A key assumption of the Poisson distribution is that the mean and variance of the process are equal; however, this is rarely the case in practice. Instead, over-dispersion often exists due to the variance being greater than the mean. In far fewer cases, such as the one presented here, data are under-dispersed (variance less than the mean). Inference concerning effects of interest can account for these over- and under-dispersed cases by relaxing the assumption of equi-dispersion and implementing a quasi-Poisson inference approach. The case study in this paper concerns the application of quasi-Poisson regression to an airborne spacing algorithm to properly account for the bias in effects tests since the more common use of the negative binomial is not applicable in the under-dispersed case (Cameron and Trivedi (1998)).

The National Aeronautics and Space Administration (NASA) conducted a computer simulation to assess the performance of a recently modified airborne spacing algorithm used in a

suite of integrated air/ground technologies that allow Interval Management (IM) operations to occur in high-density terminal environments (Swieringa et al. 2014). IM consists of flight deck automation that enables aircraft to achieve or maintain precise spacing behind a preceding aircraft, which is referred to as the target aircraft. The avionics used to conduct an IM operation include a spacing algorithm onboard the aircraft that provides commanded speeds which the flight crew follows in order to achieve or maintain the precise spacing interval.

The existing NASA spacing algorithm was modified to address integration issues with air traffic control automation that were discovered during previous human-in-the-loop simulations. A new term was added to the control law in the spacing algorithm to prevent undesirable closure rates between the IM and target aircraft, causing the IM aircraft to either get too close to the target aircraft or the aircraft behind it. Previous versions of the NASA spacing algorithm produced closure rates that were unacceptable to air traffic controllers. For more details regarding the spacing algorithm, see Swieringa et al. (2014).

A computer simulation was specifically designed to evaluate the modified algorithm before proceeding with additional, more expensive human-in-the-loop testing. Two of the key metrics used in the evaluation showed clear signs of being under-dispersed (i.e., variance less than the mean). Since inference from a Poisson regression relies upon the assumption of equi-dispersed data, resulting effects tests are biased when the data are truly over- or under-dispersed. Therefore, inference of these two key performance metrics with respect to the factors of interest in this simulation and their interactions required proper bias-adjustment to account for the under-dispersion.

Data were collected during a deterministic computer simulation with a simulated airspace that modeled the Phoenix Sky Harbor (KPHX) terminal environment. Table 1 presents the five independent variables used in this study: wind condition, target aircraft speed profile, target aircraft arrival route, initial spacing error, and expected target aircraft weight. The five wind conditions were chosen from actual winds recorded at KPHX in 2011 and correspond to wind patterns that were expected to exercise the modifications to the airborne spacing algorithm. The seven target aircraft speed profiles, created by KPHX subject matter experts, represent various trajectories flown by the target aircraft. The speed profiles were a nominal profile ("Nominal"), where the target aircraft flew the speeds expected by the spacing algorithm; a fast speed profile ("Fast"), where the target maintained a speed higher than the published speeds throughout the arrival; a slow speed profile ("Slow"), where the target aircraft maintained a speed slower than the published speed throughout the arrival; an altitude profile ("Altitude Change"), which added an altitude step-down prior to the terminal airspace to the slow speed profile; and three "pulse" speed profiles which switched between faster than and slower than the published speed profile over the course of the arrival ("Pulse 1", "Pulse 2", and "Pulse 3"). The two target aircraft arrival routes correspond to the target aircraft arriving at KPHX on the northwest (MAIER) or northeast (EAGUL) routes as depicted in Figure 1. The expected weight of the target aircraft varied from 160,000 to 198,000 pounds, and at the start of a run, the IM aircraft had an initial spacing error that was either on time, 60 seconds ahead of schedule, or 60 seconds behind schedule. This experiment utilized a factorial design, so all treatment combinations were simulated. Since the study was deterministic, only one replicate was needed, resulting in a total of 630 runs.
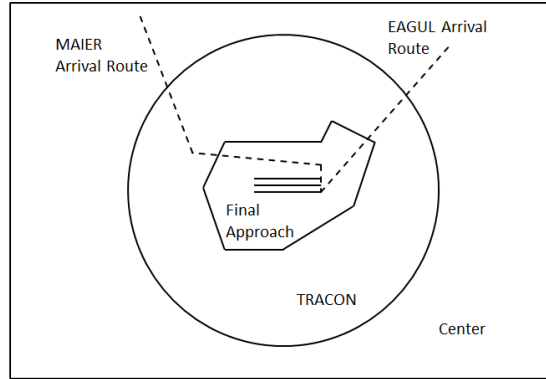
Figure 1. Depiction of the arrival routes and three regions of the airspace

Table 1. Independent variables under investigation and their levels

Independent Variable	Levels
Wind Condition (WC)	High Wind Magnitude
	Target Late and IM Aircraft Late
	Target Early and IM Aircraft Early
	Target Late and IM Aircraft Early
	Target Early and IM Aircraft Late
Target Speed Profile (TSP)	Nominal
	Altitude Change
	Fast
	Slow
	Pulse 1
	Pulse 2
	Pulse 3
Target Arrival Route (TAR)	EAGUL
	MAIER
Initial Spacing Error (ISE)	60 seconds early
	on time
	60 seconds late
Expected Target Weight (ETW)	160,000 lbs.
	185,000 lbs.
	198,000 lbs.

Before implementing the algorithm into a larger scale environment with direct human participation, researchers were interested in identifying specific conditions of the factors listed in Table 1 that degrade the spacing algorithm's performance. Two of the metrics collected during the study to quantify this behavior were spacing error inflection count and speed change rate. An

inflection occurs when the IM aircraft is more than 10 seconds behind schedule and then traverses the on-time mark to become more than 10 seconds ahead of schedule, or vice versa (see Figure 2). The number of inflections is a means by which to monitor the stability of the spacing algorithm's control law, and a large number of inflections is an indicator that the control law is underdamped. The ideal behavior of the airborne spacing algorithm is to null the spacing error and then maintain the assigned spacing interval for the duration of the flight; thus, the number of inflections is not expected to have a high dependence on the length of the flight. The speed change rate commanded by the algorithm is defined as the number of speed changes per minute during IM operations. The speed change rate was analyzed for three non-overlapping regions of the simulated airspace: Center, Terminal Radar Approach Control (TRACON), and Final Approach (see Figure 1). In practice, each of these three airspace regions have different design constraints and are controlled by different air traffic controllers using different techniques at separate facilities, and the target aircraft speed profiles in this study were designed to emulate this behavior. In addition, a single wind field varies by altitude; thus, as an aircraft descends during its flight from the Center airspace to the runway, the winds encountered vary. As a result, the correlation in the speed change rate for a single flight is considered to be negligible over these three regions of airspace. Therefore, in this study, inflection counts are analyzed per flight, while speed change rates are calculated per minute (min) for each region of the simulated airspace.
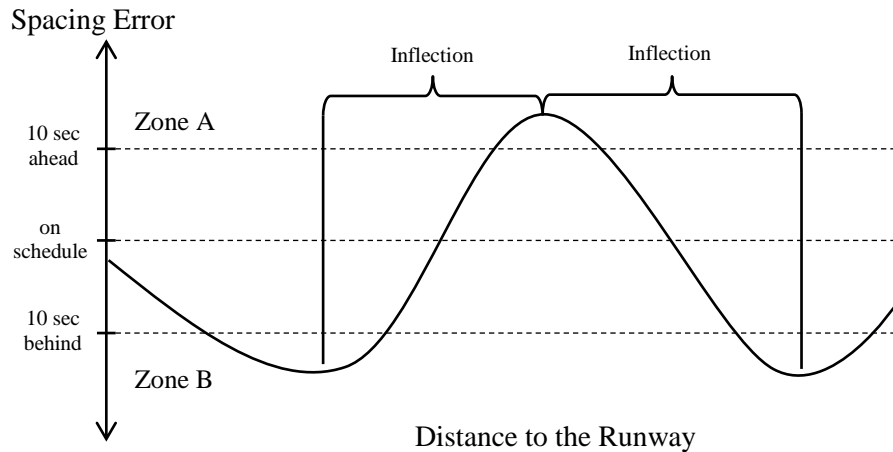

Figure 2. Depiction of two inflections occurring. First, when the IM aircraft moves from Zone B (more than 10 seconds behind schedule) to Zone A (more than 10 seconds ahead of schedule). Then, a second inflection occurs when the IM aircraft moves from Zone A back to Zone B.

METHODS

The most common method for modeling count data is implementing a Poisson regression; however, under-dispersion in all four datasets of interest from the simulation study was anticipated. That is, low variability with respect to the mean was expected since recent modifications to the algorithm included several new features intended to improve the speed control behavior by decreasing the frequency of speed changes and reducing the effect of noise on the commanded speeds. It is common in practice to encounter datasets that do not meet the equal mean/variance

4

assumption of the Poisson distribution, though under-dispersion is the rarer of the two cases. Some examples of modeling over- or under-dispersed data include the simulation study of Heinzl and Mittlbock (2003), which investigates the effect of dispersion on R-squared measures for Poisson regression models, Byers et al. (2003) where the negative binomial regression model is used for over-dispersed discrete outcomes, and Boyle et al. (1997) where under-dispersed, zero-inflated medical data were analyzed using Poisson regression models to evaluate the bias of goodness-of-fit statistics when the data sets are sparse. Unlike in Boyle et al. (1997), our data are not zero-inflated and our focus is on inference rather than on goodness-of-fit.

Poisson Regression

Consider a dataset for which the data are counts recorded within a specified unit interval (time, distance, area, etc.), i.e., rates. The Poisson distribution models the probability of y events occurring within a specified unit interval as

$$P(y|\mu) = \frac{e^{-\mu}\mu^y}{y!}, \qquad y = 0, 1, 2, \ldots$$

where μ is both the mean and variance of the distribution, and depends on the specified unit or period of time. For example, if μ is the mean number of events per unit time and t is the time period of interest, then the mean number of events in time period t is μt. This expression is based on the assumption that the mean number of events per unit is constant. However, in practice the mean often depends on levels of regressor variables that change during the process. In this case, Poisson regression can be used to model the data.

Let x_{ij} be the level of the *j-th* regressor variable at time t_i for $i = 1,\ldots, n$ and $j = 1,\ldots, k$. Then μ_i is the mean number of events in time period t_i. Assuming for $i = 1, \ldots, n$ that μ_i is not changing independently from observation to observation, then the rate $\frac{\mu}{time}$ can be modeled as a function of the k regressor variables. The Poisson mean then becomes $\mu(x_i; \boldsymbol{\beta})$ where $\boldsymbol{\beta}$ is a vector of parameters to be estimated and x_i is a vector of k regressors at time i. Using the log-link function results in the Poisson generalized linear model (GLM) also known as the log-linear regression model,

$$\frac{\mu}{time} = \mu(x_i; \boldsymbol{\beta}) = e^{x_i^T \beta} = e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}.$$

Taking the natural log of both sides,

$$\ln\left(\frac{\mu}{time}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$
$$\mu = e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + offset}$$

where the *offset* term represents the natural log of the time taken for the y events to occur. The coefficient of this offset term is fixed at one since the response is assumed to change per unit of time yet the progression of time does not affect the response. Also, when the interval of time for recording counts is the same for each observation, the offset is solely a constant that no longer aids in explaining variation from interval to interval and is therefore absorbed into the model intercept

term, β_0. We refer the reader to McCullagh and Nelder (1989, pg. 183), Myers (1990, pg. 332), and Cameron and Trivedi (1998, pg. 61) for further details concerning regression models for count data and other GLMs. For our datasets, inflection counts were analyzed as counts per simulated flight, resulting in the offset $= ln(time\ period) = ln(1\ \text{flight}) = 0$. However, all three speed change rate metrics were analyzed as the number of changes per minute, and since each flight segment varied in length of time, an offset of $ln(minutes\ in\ specified\ flight\ region)$ was added to each model.

Quasi-Poisson Modeling and Bias-Adjustment

Although Gourieroux et al. (1984) point out that parameter estimates using standard Poisson regression are consistent in the presence of over- or under-dispersion, Cameron and Trivedi (1986) show that standard errors of these estimates are biased downward in the presence of over-dispersion and upward in the presence of under-dispersion. Due to the consistency of the linear estimators, prediction is not affected by the presence of over- or under-dispersion; however, inference is directly affected by biased standard errors and their subsequent effects-test statistics. In the case of over-dispersion, this could result in a variable appearing to be significant when in fact it is not, and the opposite incorrect conclusion that a variable is not significant when in fact it is can occur in the under-dispersed case. Therefore, proper adjustment of these errors is critical for statistical inference. By adjusting the effects-test statistics, equi-dispersion is no longer an assumption and, therefore, quasi-Poisson rather than standard Poisson regression analysis should be conducted.

Cameron and Trivedi (1998) state that an alternative to Poisson regression is to specify a more general distribution than the Poisson (such as the negative binomial) that does not require equally dispersed data. Indeed, adopting the negative binomial is common practice when the data suffer from over-dispersion; however, they also note that use of the negative binomial distribution is not permissible in the case of under-dispersion since the formulation of the negative binomial requires the mean to be less than or equal to the variance. The authors go on to suggest that the easiest way to handle under-dispersed data is to conduct a standard Poisson regression to estimate model parameters since the parameter estimates are consistent and then adjust the biased standard errors of the output. Statistical software packages such as JMP and R have dispersion options available that calculate the dispersion parameter, ϕ, and automatically adjust the biased errors and effects-test statistics.

An estimate of the dispersion parameter, $\hat{\phi}$, is given by the sum of the squared standardized (Pearson) residuals divided by the residual degrees of freedom. That is,

$$\hat{\phi} = \frac{1}{n - df_{model}} \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i} \,, \tag{1}$$

where $\hat{\mu}_i = exp(x_i'\hat{\beta})$ are the fitted values from the standard Poisson regression. Here n is the number of total observations and df_{model} is the total degrees of freedom from the effects in the model (see the next section for an explanation of the degrees of freedom for an individual model effect). If adjustments for standard errors are not readily available in software, they can be easily carried out by multiplying the biased standard errors of effects (denoted by *SE*) by the square root

of the estimate of the dispersion parameter (i.e., $\sqrt{\hat{\phi}}*SE$). Test statistics are adjusted by dividing them directly by $\hat{\phi}$.

It is important to note that the dispersion parameter will always be greater than or equal to zero. Intuitively, ϕ can be interpreted by its deviation from the value of one, where $\phi < 1$ indicates under-dispersion, $\phi > 1$ indicates over-dispersion, and values of ϕ close to one indicate the standard Poisson equal mean/variance relationship holds true (i.e., equi-dispersion). In cases of perfect equi-dispersion, the sum of the squared Pearson residuals mentioned above will be equal to the residual degrees of freedom, resulting in $\phi = 1$. The sum of the Pearson residuals is also known as the Pearson χ^2 test statistic and follows an approximate χ^2 distribution with residual degrees of freedom when n is large. Therefore, deviations from equi-dispersion can be tested, though it should be noted that due to the approximate χ^2 distribution this is not an exact test and should be interpreted accordingly, especially in cases of low degrees of freedom. For over-dispersion (under-dispersion), upper (lower) quantiles are used for a specified one sided α-level hypothesis test of the Pearson χ^2 test statistic.

Inference for Poisson Regression Models

In this study, three of the five independent variables were categorical, which is very common in simulations conducted at NASA to evaluate new air traffic management technologies. Inference on Poisson regression models containing categorical factors can be conducted by testing individual "dummy" variables or by conducting Likelihood Ratio (LR) tests for each individual effect (Agresti (2007, pg. 86)). If "dummy" variables are used, a base level is chosen against which to test the other levels of the variable. LR-tests can be used to determine if the overall effect of the variable (i.e., a *combination* of *all* its levels) has a significant impact on the response. That is, LR-tests indicate whether the model fits the data better with or without the variable. Individual effects can then be tested by comparing reduced models less each variable to the full model containing all variables. Some software packages use these Type-III tests of effects as standard output. Other packages may require these Type-III test calculations to be specified beforehand or may follow effect heredity assumptions which rely on main effect and interaction parent relationships.

In the case of the equi-dispersed Poisson model, the LR-test rejects the significance of a main effect or interaction if

$$-2ln\left[\frac{L(full\ model)}{L(reduced\ model)}\right] > \chi^2_{(1-\alpha,df)} \tag{2}$$

where L is the likelihood of a model. The LR-test statistic follows a χ^2 distribution with degrees of freedom equal to one if the factor is continuous and v-1 degrees of freedom if the factor is categorical where v represents the number of levels for the tested categorical effect. Two-factor interactions where both factors are categorical can be tested in a similar fashion using the product of the degrees of freedom of the respective categorical factors. Two-factor interactions made up of one continuous and one categorical factor use degrees of freedom of the categorical factor. In the case of over- or under-dispersed data, these χ^2 statistics are biased, and thus, need to be

adjusted. This can be achieved by simply dividing each χ^2 statistic by the dispersion parameter given in (1). Therefore, for the over- or under-dispersed case, the effects-test in (2) becomes

$$-2ln\left[\frac{L(full\ model)}{L(reduced\ model)}\right] * \hat{\phi}^{-1} > \chi^2_{(1-\alpha,df)}. \tag{3}$$

The case study presented in this paper implements the Type-III LR-test because overall effects rather than level-to-level comparisons were of interest. *P*-values rather than bias-adjusted χ^2 statistics are presented to allow for easier interpretation of the results. If a practitioner is not using software to carry out the calculations, it should be noted that χ^2 statistics for effects in over- or under-dispersed cases should be calculated using likelihoods from the *standard* Poisson models before being adjusted by the dispersion parameter.

SIMULATION STUDY

Before presenting the results of the case study, we detail a small simulation study that compares standard Poisson and quasi-Poisson regression under various levels of underdispersion. The goal of the simulation study is to examine the level of underdispersion at which bias adjustments are needed. Utilizing a two-level full factorial in five factors as the design matrix (32-runs), the simulation protocol proceeds as follows for each choice of $\phi = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$ corresponding to mild to more severe underdispersion:

For each of 1,000 iterations:

1. Select *m* columns of the design matrix at random to be active. The simulation is performed for each of *m* = 2-5 active main effects.
2. Randomly assign two-factor interactions to be active. The simulation is performed for each of 1-7 active two-factor interactions.
3. Select the coefficient vector (**β**) by randomly assigning coefficients for the active effects from $\{\pm0.1, \pm0.2, ..., \pm0.9\}$. These choices are based on those observed for active effects in airborne spacing algorithm case study.
4. Compute $\boldsymbol{\mu} = e^{X_A\beta}$ where X_A is the matrix consisting of the active effect columns.
5. Generate the response vector, **y**, by simulating random variates from a Poisson distribution with mean **μ** and level of underdispersion defined by ϕ. This can be accomplished in R via the *tweedie()* function.
6. Fit a main effect and two-factor interaction model to the simulated data via *both* standard Poisson and quasi-Poisson regression and determine the set of active effects for each model fit (a significance level of α=0.05 is used).

At the end of the 1,000 iterations, the average proportion of correctly identified active effects (power) is calculated for both standard Poisson regression (denoted π_S) and quasi-Poisson regression (denoted π_Q). Figure 3 displays $\pi_S - \pi_Q$ versus ϕ across all simulation scenarios.
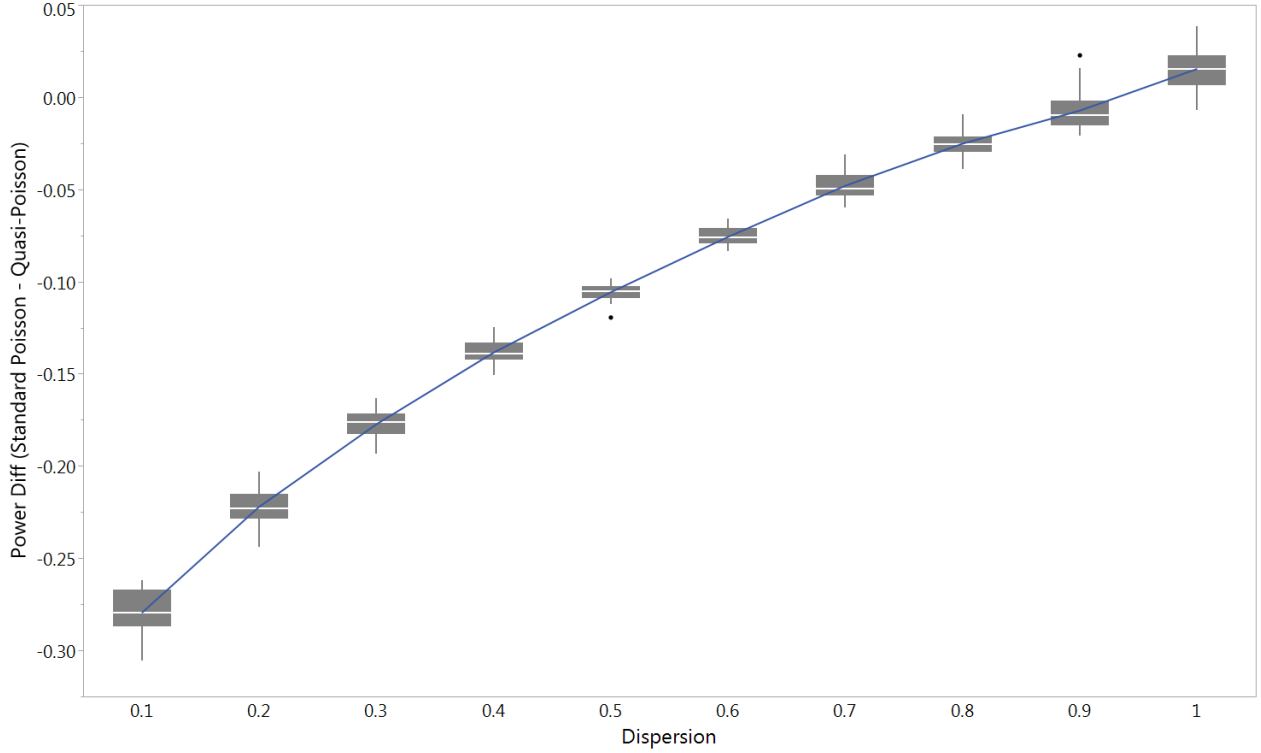
Figure 3. Simulation Results. $\pi_S - \pi_Q$ versus ϕ

By plotting $\pi_S - \pi_Q$, we can visualize the reduction in power of employing standard Poisson regression when a quasi-Poisson approach is appropriate. As the dispersion parameter approaches one, $\pi_S - \pi_Q$ approaches zero (as expected). Note that for $\phi \geq 0.6$, the reduction is power is less than 10% (which does not seem overly severe). In such cases, bias adjustments are likely unnecessary (i.e., standard Poisson regression appears robust to milder amounts of underdispersion). On the other hand, smaller values of ϕ indicate more substantial reductions in power. For instance, at $\phi = 0.1$, the mean power loss is approximately 28%. As will be seen in the next section, ϕ is estimated to be less than 0.5 for each of the case study responses. Given the potential for loss of power, bias adjustments are appropriate.

CASE STUDY RESULTS

Though under-dispersion was anticipated for the simulation study presented in this paper, the researchers were interested in testing for the significance of under-dispersion before making adjustments. From Table 2 we can see that dispersion parameter values for each of our datasets are less than one, indicating proper bias-adjustment would likely show some effects as significant even though standard Poisson regression analysis may not. Therefore Pearson χ^2 tests were carried out and revealed significant under-dispersion in each of our four datasets (all p-values < 0.0001).

Table 2. Dispersion parameters for inflection count and speed change rates

Metric	$\hat{\phi}$	p-value for Pearson χ^2 Test for Under-Dispersion
Inflection Count	0.43	< 0.0001
Speed Change Rate for Center	0.33	< 0.0001
Speed Change Rate for TRACON	0.25	< 0.0001
Speed Change Rate for Final Approach	0.22	< 0.0001

Both the non-adjusted and bias-adjusted Type-III tests of effects for inflection count and speed change rate are shown in Table 3. In both cases for inflection count, none of the main effects are significant and the same six two-factor interactions involving wind condition, target aircraft arrival route, target aircraft speed profile, and initial spacing error are significant at the $\alpha = 0.05$ level; however, note the smaller p-values as a result of properly adjusting for under-dispersion. The last six columns of Table 3 show the change in p-values when adjusting for under-dispersion in each of the three speed change rate models. For the Center region, the quasi-Poisson effects tests indicate two significant interaction effects not found by the standard Poisson tests. For the TRACON and Final Approach regions, the standard analysis results in no statistically significant effects, while the bias-adjusted analysis detected a number of statistically significant main effects and two-factor interactions.

Table 3. Comparison of Type-III tests of effects when using standard non-adjusted results and bias-adjusted results

Effect	Inflection Count		Speed Change Rate					
			Center		TRACON		Final Approach	
	Standard P-value	Quasi P-value	Standard P-value	Quasi P-value	Standard P-value	Quasi P-value	Standard P-value	Quasi P-value
WC	0.9702	0.8731	0.8599	0.4143	0.4948	0.0082 *	0.9093	0.3230
TAR	0.8740	0.8096	0.5975	0.3596	0.9771	0.9540	0.8894	0.7642
TSP	0.9811	0.8614	0.9940	0.9024	0.7326	0.0243 *	0.9399	0.2228
ETW	0.6262	0.4593	0.9060	0.8376	0.6263	0.3270	0.7046	0.4135
ISE	0.3881	0.1899	0.6164	0.3847	0.7923	0.5961	0.8164	0.6165
WC*TAR	< 0.0001 *	< 0.0001 *	0.0010 *	< 0.0001*	0.1173	< 0.0001 *	0.2184	< 0.0001 *
WC*TSP	< 0.0001 *	< 0.0001 *	0.0330 *	< 0.0001*	0.5016	< 0.0001 *	0.5284	< 0.0001 *
WC*ETW	0.9530	0.8117	0.9773	0.8467	0.7260	0.0806	0.9120	0.3328
WC*ISE	< 0.0001 *	< 0.0001 *	< 0.0001 *	< 0.0001*	0.8702	0.2817	0.9938	0.8981
TAR*TSP	< 0.0001 *	< 0.0001 *	0.0044 *	< 0.0001*	0.3338	0.0001 *	0.2927	< 0.0001 *
TAR*ETW	0.8700	0.8037	0.8609	0.7611	0.7738	0.5629	0.9462	0.8842
TAR*ISE	< 0.0001 *	< 0.0001 *	0.0866	0.0030*	0.0898	0.0006 *	0.5687	0.2189
TSP*ETW	0.9645	0.7726	0.9929	0.8892	0.7069	0.0181 *	0.9809	0.5208
TSP*ISE	< 0.0001 *	< 0.0001 *	0.3868	0.0041*	0.6918	0.0151 *	0.7498	0.0133 *
ETW*ISE	0.7771	0.6672	0.4468	0.1869	0.9584	0.9163	0.8737	0.7317

An example of the benefit of adjusting for under-dispersion can be seen in Figure 4. The first figure shows that the high magnitude wind condition produced the largest speed change rate in the TRACON when combined with the altitude change, pulse 2, pulse 3, and slow speed profiles. The second figure shows that on Final Approach, the highest speed change rate occurred with the

target late / IM early and target early / IM late wind conditions combined with the fast and nominal target aircraft speed profiles. However, the interaction between the wind condition and the target aircraft speed profile for speed change rate in the TRACON and on Final Approach was only found to be statistically significant after the quasi-Poisson regression was adjusted for under dispersion. The statistical significance of the results indicate that an increase in the speed change rate degrades algorithm performance since it can increase pilot workload as well as decrease the efficiency of the aircraft. These results may have been missed if the quasi-Poisson regression had not been adjusted for under-dispersion.

Interaction effects between wind condition, target aircraft speed profile, target aircraft arrival route, and initial spacing error were found to be significant for both inflection count and speed change rate in this initial computer simulation. These independent variables were further investigated in a follow-on study which evaluated performance at multiple airports. None of the effects associated with expected target aircraft weight were significant, indicating that the algorithm is robust to target aircraft weight. Since changes in the weight of the target aircraft does not affect performance, this independent variable did not need to be further investigated. These findings can be used to better understand the conditions that affect the performance of the algorithm and provide valuable information for future algorithm improvements.
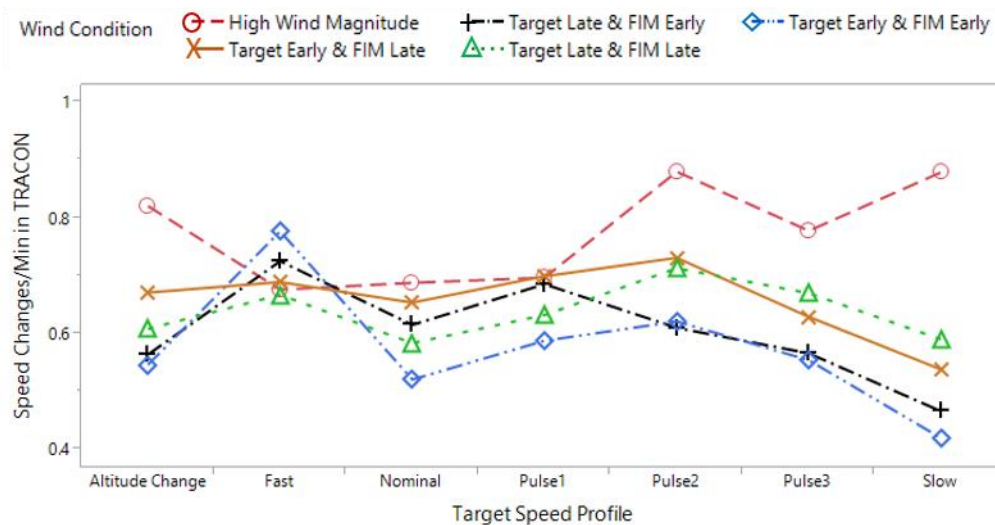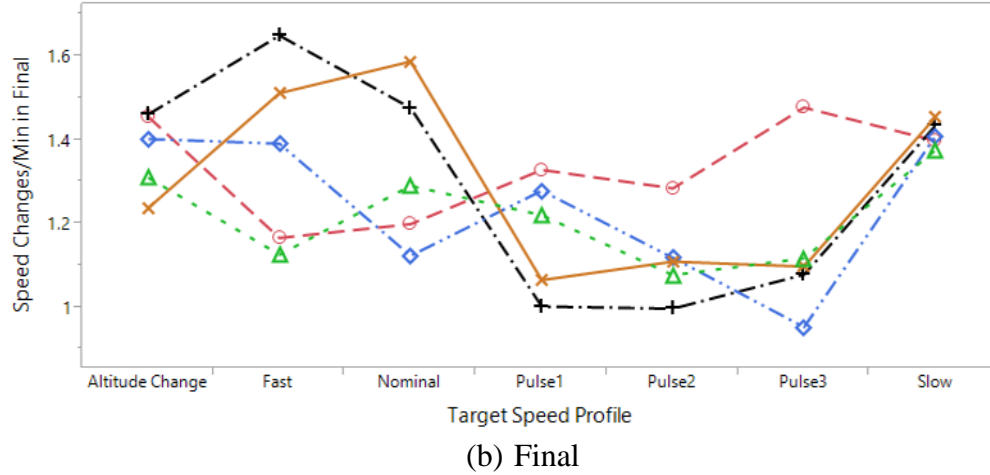

(a) TRACON

(b) Final

Figure 4. Interaction plots of the target aircraft speed profile by wind condition for speed change rate in the (a) TRACON and on (b) Final

Comments on Statistical Software Packages

Not all statistical software packages appear to fully support the ability to provide dispersion tests for both under- and over-dispersion cases. As an example, JMP Pro 13 will produce the Pearson χ^2 goodness-of-fit statistic; however, the subsequent test result reflected in a *p*-value is only for over-dispersion. Hence, for cases of significant under-dispersion, this software will correctly indicate no over-dispersion, but a test for under-dispersion as outlined above will not be given. That is, the software only provides a test for over-dispersion by implementing upper quantiles for a specified one sided *α*-level hypothesis test of the Pearson χ^2 test statistic. For our cases, we were interested in testing for significant under-dispersion. Therefore, lower quantiles were instead used and separate calculations using the R statistical software package revealed significant under-dispersion in each of our four datasets (see Table 2).

Concerning inference, JMP and R statistical packages support quasi-Poisson analysis for both under- and over-dispersed data. In JMP Pro 13, a dispersion option in the GLM model fitting tool can be selected prior to fitting the log-linear model. With this option, Type-III tests of effects are provided and likelihood ratio χ^2 statistics are automatically bias-adjusted by dividing each by ϕ. Although the overall dispersion test results from this software package are always in terms of over-dispersion, effects tests are still bias-adjusted in the proper direction for either the under- or over-dispersed case. The R statistical software package can also accommodate quasi-Poisson analysis through the use of the *glm()* function within the *stats* package with the *family=quasipoisson* option, which provides a dispersion parameter as standard output. Type-III tests of effects can be calculated in R using the *Anova()* function in the *cars* package (see appendix).

CONCLUSIONS

A summary of the statistically significant factors when adjusting for under-dispersion is shown in Table 4. The high magnitude wind condition produced the largest speed change rate in

the TRACON when combined with the altitude change, pulse 2, pulse 3, and slow target aircraft speed profiles. On Final Approach, the highest speed change rate occurred with the target late / IM early and target early / IM late wind conditions combined with the fast and nominal target aircraft speed profiles. These results can be used to better understand performance of the algorithm and provide valuable information for future algorithm improvements. For example, the performance of the algorithm could be improved in the future if more detailed wind information was available to the IM aircraft. The true impact of this result would have been mistakenly identified as insignificant had the Poisson distribution assumption of equal mean and variance been enforced.

It can be seen in Table 4 that the main effects are rarely statistically significant for any of the metrics even after being bias-adjusted, while two-factor interactions between wind condition, target aircraft arrival route, target aircraft speed profile, and initial spacing error are almost always statistically significant. The dynamics of the system, including the algorithm and aircraft dynamics, are very complex. This indicates that changes in a single input to the system may have very little effect on the system output; however, changes in multiple factors can create significant changes in the output. Ideally, the algorithm's performance would be robust to external factors, so that performance would be consistent under all conditions. Detection of the statistically significant effects identified specific conditions that degrade algorithm performance. These factors were further evaluated in a follow-up computer simulation prior to additional expensive human-in-the-loop testing. Based on the results of the simulation presented in this paper, the follow-up study also examined a modified process for providing wind information to the algorithm. By not properly taking into account the under-dispersion present in the data, statistical inference would deem most if not all of the effects to be non-significant in a study where actual relationships between the factors and responses deserve closer attention. Though not the case in the study presented here, just as important is the detection of over-dispersion and the ability to test if its influence requires tests of effects to be bias-adjusted upward. The same quasi-Poisson methodology can be implemented to account for over-dispersion as well as under-dispersion.

Table 4. Summary of significant effects for all metrics investigated when the bias due to under-dispersion is properly adjusted

Effect	Inflection Count	Speed Change Rate		
		Center	**TRACON**	**Final Approach**
WC			X	
TAR				
TSP			X	
ETW				
ISE				
WC*TAR	X	X	X	X
WC*TSP	X	X	X	X
WC*ETW				
WC*ISE	X	X		
TAR*TSP	X	X	X	X
TAR*ETW				

TAR*ISE	X	X	X	
TSP*ETW			X	
TSP*ISE	X	X	X	X
ETW*ISE				

APPENDIX: R CODE

Quasi-Poisson regression can be carried out in a straightforward manner in R. The following R code was used for the Inflection Count response:

```
options(contrasts=c(unordered="contr.sum",ordered="contr.poly"))

#Fit a main effects and two-factor interaction model using
quasi-Poisson regression
myglm <-
glm(InflectionCount~(Wind.Condition+TgtArrivalRoute+Target.Speed
.Profile+ExpectedTgtWeight+InitialSpacingError..sec.)^2,family="
quasipoisson",data=mydata)

#The car package is needed for requesting tests based on Type
III sums of squares
library(car)
TypeIII_SS <- Anova(myglm,type="III")
```

ACKNOWLEDGEMENTS

REFERENCES

Agresti, A. (2007). *An Introduction to Categorical Data Analysis, 2nd ed*. New York, NY: John Wiley and Sons.

Baxley, B.T., Johnson W.C., Swenson H.N., Robinson J.E., Prevot T., Callantine T.J., Scardina J., and Greene M. (2013). "Air Traffic Management Technology Demonstration-1 Concept of Operations (ATD-1 ConOps), Version 2.0". *NASA/TM-2013-218040*.

Boyle, P., Flowerdew, R., and Williams, A. (1997). "Evaluating the Goodness of Fit in Models of Sparse Medical Data: A Simulation Approach". *International Journal of Epidemiology,* 26:651-656.

Byers, A.L., Allore, H.; Gill, T.M., and Peduzzi, P.N. (2003). "Application of negative binomial modeling for discrete outcomes: A case study in aging research". *Journal of Clinical Epidemiology,* 56:559-564.

Cameron, A.C., and Trivedi, P.V. (1986). "Econometric models based on count data: Comparisons and applications of some estimators and tests". *Journal of Applied Econometrics,* 1:29-53.

Cameron, A.C., and Trivedi, P.V. (1998). *Regression Analysis of Count Data*. Cambridge University Press, New York, NY.

Gourieroux, C., Monfort, A., and Trognon, A. (1984). "Pseudo maximum likelihood methods: Applications to Poisson models". *Econometrica*, 52(3):701-720.

Heinzl, H., and Mittlbock, M. (2003). "Pseudo R-squared measures for Poisson regression models with over- or under-dispersion". *Computational Statistics & Data Analysis,* 44:253-271.

McCullagh, P., and Nelder, J.A. (1989). *Generalized Linear Models, 2nd ed*. New York, NY: Chapman and Hall.

Myers, R.H. (1990). *Classical and Modern Regression with Applications, 2nd ed*. Belmont, CA: Duxbury Press.

Swieringa, K.A., Underwood, M.C., Barmore, B., and Leonard, R.D. (2014). "An evaluation of a flight deck interval management algorithm including delayed target trajectories". AIAA Aviation Technology, Integration, and Operations Conference, AIAA Paper 2014-3148, Atlanta, GA, June pp. 16-20.